



Indian Institute of Technology Kanpur

Course IME 692, Fall (Odd) Semester 2017

Time and Place: M 17:00-18:30 C3;
T 10:00-11:30 C3

About IIT Kanpur

Indian Institute of Technology Kanpur carries out original research of significance and technology development at the cutting edge. It imparts training for students to make them competent, motivated engineers and scientists. The Institute not only celebrates freedom of thought, cultivates vision and encourages growth, but also inculcates human values and concern for the environment and the society.

Course: IME 692, Advanced Statistical Methods for Business Analytics

Instructor Name: Shankar Prawesh

Contact Information: IME 307

Phone: +91-512-259-6182

Email: sprawesh@iitk.ac.in

Faculty Office Hours: Monday 16:00-17:00, Tuesday 11:30 am-12:30 pm (IME 307)

Course Description and Objectives

The course is designed to train students on understanding research problems and situations requiring multivariate approaches, selecting appropriate multivariate techniques of data analysis, interpreting the results of analysis, and applying the techniques to business and research problems. The course includes topics dealing with multiple interdependent techniques (such as Principal Component Analysis, Factor Analysis, Cluster Analysis, Multidimensional Scaling), multiple dependent techniques (such as Multiple Regression Analysis, Discriminant Analysis, Path analysis, Design of Experiments and Conjoint Analysis) and recent advances in dealing with Big Data.

Students are also required to read and review research papers related to the application of statistical methods throughout the semester. It is expected that this exercise will help students who are working towards their Masters' thesis or Doctoral dissertation.

Prerequisite

Probability and statistics (IME 603 & MBA 633), exposure to R.

Course Materials, Text etc.

Course Material: Power point slides if used for lectures will be provided to students. But students should recognize the slides are mnemonics for the instructor, and in no way substitute for lecture notes and text material. Hence students are **STRONGLY ADVISED** to take class notes.

The following texts will be used for reference.

Text/Reading Materials:

1. Applied Multivariate Statistical Analysis (by Wolfgang Härdle & Léopold Simar), 2nd edition, Springer, 2007. (soft copy available from library)
2. An Introduction to Statistical Learning *with Applications in R* (by Gareth James, Witten Daniela, Hastie Trevor, & Tibshirani Robert), Springer, 2014.

Soft copy of the book: <http://www-bcf.usc.edu/~gareth/ISL/>

Teaching method and course emphasis: The course will be mostly based on lectures, research papers and presentations. Apart from in class learning, away from class learning will be encouraged through assignments. Other material/readings will be handed out from time to time.

Course Policies

All home works are with fixed deadlines. These assignments are designed to provide you with a chance to put the concepts introduced in the class into practice.

Attendance Policy The attendance of students in the class will be monitored regularly, and it will be considered during their evaluation of class participation. Class attendance and participation is expected, and has been proven as a prerequisite for doing well. Students are expected to be prepared for class and able to contribute to the class discussion. Students may be called on at random to offer their views. If you miss a class, it is your responsibility to cover the material.

Cell phone Smart phone, iPhones, iPads, MP3 players are to be turned off during class.

Food and drink No food inside the classroom. Only a water bottle is permitted (No other beverages).

Use of laptops/tablets Empirical evidence of: multitasking using laptop in classroom hinders the learning experience. Therefore, no emails/chatting/social-media are allowed during the class.

Course Grade Determination

Exams and grading policy Mid-semester exam 25, End-semester exam 40, Quiz 15, Take-home assignments* 10, Class participation 10

*In a group of two

Grading policy

Total points (TP) ≥ 85 - A

$85 > TP \geq 75$ - B

75 > TP >= 65 - C

65 > TP >= 55 - D

55 > TP >= 45 - E

45 > TP – F

Resource Guide

Official Class Cancellation

In the event that the Institute officially cancels classes due to some reason, the class activity (homework, test, etc.) scheduled for the canceled date is automatically rescheduled for the next date that class is officially in session.

Honor System: Plagiarism and Academic Integrity

The honor system assumes the responsibilities of students and faculty in upholding academic integrity, while at the same time respecting the rights of individuals to the due process offered by administrative hearings and appeals. Accordingly, “members of this academic course are required to conduct themselves in accordance with the highest standards of academic honesty and integrity.” In addition, all members are required to:

- Agree to be bound by the Honor System and its procedures;
- Report suspicion or knowledge of possible violations of the Honor System;
- Support an environment that reflects a commitment to academic integrity;
- Answer truthfully when called upon to do so regarding Honor System cases, and,
- Maintain confidentiality regarding specific information in Honor System cases.

Most importantly, “All students are presumed upon enrollment to have acquainted themselves with and have an understanding of the Honor System”.

In this class, because coursework may be collaborative at times, particular issues of integrity arise. You should not copy or print another student’s work without permission. Any material (this includes IDEAS and LANGUAGE) from another source must be credited, whether that material is quoted directly, summarized, or paraphrased. **In other words, you should respect the work of others and in no way present it as your own.**

Student with Disabilities

If you have a physical or mental impairment that requires an academic adjustment or accommodation, arrange a meeting with me at your earliest convenience.

Student Conduct in the Classroom The instructional program is based upon the premise that students enrolled in a class are entitled to receive instruction free from interference by other students. Accordingly, in classrooms, laboratories, studies, and other learning areas, students are expected to conduct themselves in an orderly and cooperative manner so that the faculty member can proceed

with the customary instruction. Faculty members (including graduate teaching assistants) may set reasonable standards for classroom behavior in order to serve these objectives.

Religious Observances Students desiring to observe religious holiday of special importance must provide advance written notification to the instructor by the end of the second week of classes.

Email Policy e-mail is the most convenient way to contact the instructor.

Important Dates

<http://www.iitk.ac.in/doaa/data/calendar2017.pdf>

Tentative Schedule

Schedule of Topics (Subject to Change)

Outline of Topics and Course Schedule Because of unforeseen circumstances and responsibilities it is possible the instructor will not be able to meet the scheduled class period. In case this event occurs, barring an emergency, the class will be notified and the class period will be rescheduled. So, be advised that the following class schedule is *tentative* and subject to alteration as dictated by the pace, ability and interest of the students. In class discussions, material difficulty and external forces may also dictate alterations. **Important: Sequence of topics may change.**

Class 1	Discussion of syllabus, A brief overview of course, Introduction, Basics of R
Class 2-3	Multivariate Distributions
Class 4-7	Basics of Multivariate Normal Distribution
Class 8	Principal Component Analysis
Class 9-10	Factor Analysis
Class 11-12	Clustering Techniques, Multidimensional scaling
Class 13-14	Choice modeling
Class 15-16	Bias-variance tradeoff, Bayes classifier, K-NN, Discriminant Analysis
Class 17-20	Resampling Methods: Cross-Validation, Bootstrap
Class 21-23	High-dimensional regression techniques, LASSO, LAR
Class 24-28	Tree-based Methods: Decision Trees, Regression trees, Bagging, Boosting, Random Forests

The course will also consist of reading and reviewing journal articles at regular intervals. Following are the list of articles¹:

Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, 80(6), 97–121.

Cao, L. (2017). Data science: a comprehensive overview. *ACM Computing Surveys (CSUR)*, 50(3), 43.

Einav, L., & Levin, J. (2014a). Economics in the age of big data. *Science*, 346(6210), 1243089.

Einav, L., & Levin, J. (2014b). The data revolution and economic analysis. *Innovation Policy and the Economy*, 14(1), 1–24.

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research. *Academy of Management Journal*, 59(5), 1493–1507.

Lazer, D., & Radford, J. (2017). Data ex Machina: Introduction to Big Data. *Annual Review of Sociology*, (0). Retrieved from <http://www.annualreviews.org/doi/abs/10.1146/annurev-soc-060116-053457>

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4), 1187–1230.

Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of p-values. *Royal Society Open Science*, 1(3), 140216.

Kan, A. (2017). Machine learning applications in cell image analysis. *Immunology and Cell Biology*, 95(6), 525–530.

¹ The list is tentative.